

MATRIX REVOLUTIONS THE ORIGIN OF QUANTUM VARIABLES

EDWARD G. EFFROS
DEDICATED TO RICHARD V. KADISON AND MASAMICHI TAKESAKI

FOR TRANSMITTING VON NEUMANN'S VISION
EXPANDED LECTURE NOTES

FRONTIERS IN MATHEMATICS
TEXAS A&M, NOVEMBER, 2003

ABSTRACT. This is an informal guide to the history of Heisenberg's matrix mechanics. It is designed for mathematicians with only a minimal background in either physics or geometry, and it is based upon Heisenberg's original arguments.

1. INTRODUCTION

The most dramatic shift in Twentieth Century physics stemmed from Heisenberg's formulation of matrix mechanics [9]. In classical physics, quantities such as position, momentum, and energy are regarded as *functions*. In quantum theory one replaces the functions by non-commuting *infinite matrices* or to be more precise, self-adjoint operators on Hilbert spaces. This enigmatic step remains the most daunting obstacle for those who wish to understand the subject.

Although there exist many excellent mathematical introductions to quantum mechanics (see, e.g., [12], [17]) they are understandably focused on the development of mathematically coherent methods. As a result, mathematics students must postpone understanding *why* non-commuting variables appeared in the first place. To remedy this, one can adopt a more historical approach, such as that found in Emch's beautiful historical monograph [6], the entertaining yet informative "comic book" [11], or Born's classic text [2].

In recounting the creation of quantum mechanics, the most difficult task is to describe how Heisenberg found the canonical commutation relation

$$(1) \quad PQ - QP = \frac{h}{2\pi i} I$$

for the position and momentum operators Q and P . This equation is the final refinement of Planck's principle that a certain action variable is discrete, or more precisely that it can assume only integer increments of a universal constant h . Heisenberg used the more sophisticated formulation of Bohr and Sommerfeld that

Date: January 29, 2004.

The author was partially supported by an NSF grant.

for periodic systems one has the “quantum condition”

$$(2) \quad \oint pdq = nh$$

(see §4).

In the words of Emch (see [6], p. 262) “one can only propose some very loose a priori justifications” for the derivation of (1) from (2). Even Born, who was apparently the first to postulate the general form of (1) (see [6], p. 264), avoided discussing it, appealing instead to the Schrödinger model ([2], p.130, see also [11] p.224), and this is the approach that one finds in most physics texts. We will attempt to make Heisenberg’s direct conceptual leap a little less mysterious, by deciphering an argument that Heisenberg presented in his 1930 survey [10]. At the heart of his computation is the observation that

*the analogue of the derivative for the discrete action variable is just
the corresponding finite difference quotient.*

(see (13)).

Shortly after Heisenberg introduced matrix mechanics, Schrödinger found an alternative quantum theory based upon the study of certain wave equations [16]. His approach enabled one to avoid a direct reference to Heisenberg’s matrices. Although it is both intuitive and computationally powerful, “wave mechanics” is not as useful in quantum field theory. The difficulty is that it does not fully accommodate the particle aspects of quanta. In quantum field theory one must take into account the incessant creation and annihilation of particles associated with the relativistic equivalence of mass and energy. In particular, the number of particles present must itself be regarded as an integer valued quantum variable. In Born’s words [2], p. 130, “Heisenberg’s method turns out to be more fundamental.”

Our goal has been to maximize the accessibility of the material. In order to do this we have taken liberties with the mathematical, physical, and historical details. To some extent this is justified by the fact that regardless of how much care we might take, the discussion is necessarily tentative. Although Heisenberg’s argument is mathematically quite suggestive, in the end we must discard these notions in favor of the operator techniques that grew out of them.

2. ATOMIC SPECTRA, FOURIER SERIES AND MATRICES

The crisis that occurred in classical physics is clearly seen in the peculiar properties of atomic spectra. If one sends an electric discharge through an elemental gas A such as hydrogen or sodium, the gas will emit light composed of very precise (angular) frequencies ω . The corresponding *spectrum* $\text{sp}A$ of such frequencies is quite specific to the element A . For a single frequency we have the corresponding representation

$$\cos(\omega t + a) = \text{Re } e^{i(\omega t + a)} = c_{-1}e^{-i\omega t} + c_0 + c_1e^{i\omega t}$$

for suitable complex constants c_{-1} , c_0 , c_1 . Superimposing these frequencies, we may describe the radiation by the sum

$$(3) \quad f_A(t) = \sum_{\omega \in \text{sp}_0 A} c_\omega e^{i\omega t},$$

where $\text{sp}_0 A = \text{sp}A \cup -\text{sp}A \cup \{0\}$.

There are obvious classical analogues of this phenomenon. If one strikes an object, the resulting sound can be decomposed into certain specific angular frequencies. In the case of a tuning fork, the resulting motion is harmonic, and one obtains a corresponding Fourier series for the amplitude of the sound wave in the form

$$f(t) = A \cos(\omega t + a) = c_{-1} e^{-i\omega t} + c_0 + c_1 e^{i\omega t}.$$

for suitable complex coefficients c_k . If one instead plucks a guitar string, the resulting sound is a combination of various frequencies, all of which are overtones, i.e., multiples of a fundamental frequency ω . Thus one has a Fourier series

$$(4) \quad f(t) = \sum_{n \in \mathbb{Z}} c_n e^{i(n\omega)t}.$$

where for simplicity we assume that only finitely many of the c_n are non-zero. We define the (full) *spectrum* of f to be the cyclic group $\mathbb{Z}\omega = \{n\omega : n \in \mathbb{Z}\}$. As is well-known, one can duplicate the sound of a guitar string by superimposing the pure frequencies as in (4).

More complicated systems (such as a bell) will have several degrees of freedom, and thus several fundamental frequencies. For a system with two degrees of freedom one will have two fundamental frequencies ω, ω' with the “almost periodic” expansions

$$f(t) = \sum_{n, n' \in \mathbb{Z}} c_{n, n'} e^{i(n\omega + n'\omega')t}.$$

We will restrict our attention to one degree of freedom.

The linear space $\mathcal{A}(\omega)$ of all functions of the form (4) with finitely many non-zero terms is closed under multiplication since if we are given

$$\begin{aligned} f(t) &= \sum_{n \in \mathbb{Z}} c_n e^{i(n\omega)t} \\ g(t) &= \sum_{n \in \mathbb{Z}} d_n e^{i(n\omega)t} \end{aligned}$$

then

$$f(t)g(t) = \sum_{k, n \in \mathbb{Z}} c_k d_{n-k} e^{i(k\omega + ((n-k)\omega)t} = \sum_{n \in \mathbb{Z}} a_n e^{i(n\omega)t},$$

where a_n is the “convolution”

$$(5) \quad a_n = (c * d)_n = \sum_{k \in \mathbb{Z}} c_k d_{n-k},$$

Furthermore $\mathcal{A}(\omega)$ is closed under conjugation since

$$\bar{f}(t) = \sum c_n^* e^{i(n\omega)t}$$

where $c_n^* = \overline{c_{-n}}$. In more technical terms, the $*$ -algebra $\mathcal{A}(\omega)$ is a representation of the group $*$ -algebra $\mathbb{C}[\mathbb{Z}]$. This result, of course, stems from the fact that $\text{spf} = \mathbb{Z}\omega$ is a group under addition.

Returning to atomic spectra, it is tempting to regard (3) as some kind of Fourier series. There are several problems with this interpretation.

First of all, we are actually interested in analyzing the property of a *single atom*. In this case it is inappropriate to actually “add up” the series (3). For example (getting a little ahead of ourselves) a hydrogen atom will radiate only one frequency

at a time corresponding to the electron taking a particular orbital jump. Thus superpositions do not occur when one “watches” a single atom. For this reason it is more accurate to let $f(t)$ stand for the *array* $(c_\omega e^{i\omega t})_{\omega \in \text{sp}_0 A}$.

Secondly, in striking contrast to the classical models, it is not useful to consider the additive group generated by $\text{sp}_0 A$. Given $\omega \in \text{sp}_0 A$, one cannot expect to find *any* of the overtones $n\omega$ in $\text{sp}_0 A$. Nevertheless the set $\text{sp}_0 A$ does display an exquisitely precise algebraic structure, called the *Ritz combination principle*. We may doubly index $\text{sp}A$, i.e., we may let $\text{sp}_0 A = \{\omega_{m,n}\}_{m,n \in \mathbb{N}}$, in such a manner that

$$(6) \quad \omega_{m,n} + \omega_{n,p} = \omega_{m,p}$$

for all $m, n, p \in \mathbb{N}$. In particular, $\omega_{m,m} + \omega_{m,m} = \omega_{m,m}$ and thus $\omega_{m,m} = 0$. Furthermore, $\omega_{m,n} + \omega_{n,m} = \omega_{m,m} = 0$, and therefore $\omega_{n,m} = -\omega_{m,n}$. Using this double indexing of the spectrum, our array becomes a *matrix*:

$$(7) \quad f(t) = [a_{m,n} e^{i\omega_{m,n} t}]_{m,n \in \mathbb{N}}.$$

The set $\mathcal{M}(\omega)$ of matrices (7) is already a linear space. Owing to (6), $\mathcal{M}(\omega)$ is closed under matrix multiplication and the adjoint operation since

$$\begin{aligned} f(t)g(t) &= \left[\sum_k c_{m,k} e^{i\omega_{m,k} t} d_{k,n} e^{i\omega_{k,n} t} \right] \\ &= \left[\sum_k c_{m,k} d_{k,n} e^{i(\omega_{m,k} + \omega_{k,n}) t} \right] \\ &= \left[\sum a_{m,n} e^{i\omega_{m,n} t} \right] \end{aligned}$$

where $a = cd$ is the usual matrix product, and

$$f(t)^* = [\bar{a}_{n,m} e^{-i\omega_{n,m} t}] = [a_{m,n}^* e^{i\omega_{m,n} t}]$$

with a^* the adjoint matrix. In fact one can regard $\mathcal{M}(\omega)$ as a representation of the $*$ -algebra $\mathbb{C}[\mathbb{N} \times \mathbb{N}]$ of the full groupoid $\mathbb{N} \times \mathbb{N}$. This point of view has been explored by Connes [3], but we will not pursue it further in this paper.

It is easy to prove that any doubly indexed family $\omega_{m,n}$ satisfying (6) must have the form

$$\omega_{m,n} = C_m - C_n$$

for suitable constants C_m . The precise values for the hydrogen atom are given by Balmer’s equation

$$(8) \quad \omega_{m,n} = 2\pi R \frac{c}{m^2} - 2\pi R \frac{c}{n^2}$$

where c is the speed of light, and R is known as the Rydberg’s constant.

Long before matrices were introduced, Bohr justified Rydberg’s equation by combining Rutherford’s model of the atom with a quantum condition on the action variable. This “old” quantum theory was to play a crucial role in the evolution of matrix mechanics.

3. ACTION AND QUANTIZATION CONDITIONS

Action is perhaps the least intuitive of the standard notions of classical mechanics. As usual, the easiest way to understand a physical quantity is to consider its units or “dimensions”. We let \mathcal{M} , \mathcal{L} , and \mathcal{T} denote units of mass m , length (or position) q , and time t (e.g. one can use grams, meters, and seconds). Given a physical quantity P , we let $[P]$ denote its units. We have, for example,

$$\begin{aligned} [\text{velocity } v] &= \left[\frac{dq}{dt} \right] = \mathcal{L}\mathcal{T}^{-1} \\ [\text{acceleration } a] &= \left[\frac{d^2q}{dt^2} \right] = \mathcal{L}\mathcal{T}^{-2} \\ [\text{momentum } p] &= [mv] = \mathcal{M}\mathcal{L}\mathcal{T}^{-1} \\ [\text{force } F] &= [ma] = \mathcal{M}\mathcal{L}\mathcal{T}^{-2} \\ [\text{potential energy } V] &= [-Fq] = \mathcal{M}\mathcal{L}^2\mathcal{T}^{-2} \\ [\text{kinetic energy } T] &= \left[\frac{1}{2}mv^2 \right] = \mathcal{M}\mathcal{L}^2\mathcal{T}^{-2} \\ [\text{total energy } H] &= [V + T] = \mathcal{M}\mathcal{L}^2\mathcal{T}^{-2} \end{aligned}$$

Noting that they have the same dimensions, we simply regard V , T , and $E = V + T$ as “different forms” of energy. We will often consider derivative and integral versions of these quantities, such as v and a above and the potential energy

$$V = - \int F(q) dq.$$

The dimensions frequently mirror physical laws. For example the equation for force corresponds to Newton’s second law. On the other hand the relativistic equation $E = mc^2$ corresponds to $\mathcal{M}\mathcal{L}^2\mathcal{T}^{-2} = \mathcal{M} \times (\mathcal{L}\mathcal{T}^{-1})^2$.

The usual form of a travelling wave (in one spatial dimension) is given by

$$(9) \quad f(t, q) = A \cos(\omega t + kq)$$

where ω is the angular frequency (radians per second) and k is angular wavenumber (radians per meter). The corresponding dimensions are

$$\begin{aligned} [\text{angular frequency } \omega] &= [\text{radians}]/[\text{time}] = \mathcal{T}^{-1}. \\ [\text{angular wavenumber } k] &= [\text{radians}]/[\text{distance}] = \mathcal{L}^{-1} \end{aligned}$$

We recall these are related to the frequency ν (cycles per second) and wavelength (of a cycle) λ by $\omega = 2\pi\nu$ and $k = 2\pi/\lambda$.

Given an angular co-ordinate θ measured in radians, we have the dimension

$$[\text{angular velocity } \omega] = \left[\frac{d\theta}{dt} \right] = \mathcal{T}^{-1}$$

By analogy with the momentum formula $p = mv$, the angular momentum is defined by $L = \iota\omega$, where ι is the “moment of inertia”, or equivalently L is the signed length of the vector $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, where \mathbf{r} is the position vector and \mathbf{p} is the momentum vector. Thus we have

$$[\text{angular momentum } L] = \mathcal{M}\mathcal{L}^2\mathcal{T}^{-1}.$$

In classical physics, the (restricted) action along a parametrized curve γ is defined by the formulas

$$J[\gamma] = \int_{\gamma} p dq = \int_a^b T dt,$$

and the actual motion taken by the particle is determined by finding the stationary values of suitable variations of J with fixed energy (alternatively one can use a different variational principle involving the Lagrangian, see [7], [8]). The corresponding dimensions are given by

$$[\text{action } J] = [\text{energy}] \times [\text{time}] = [\text{momentum}] \times [\text{distance}] = \mathcal{ML}^2\mathcal{T}^{-1}.$$

We see from above that action has the same dimensions as angular momentum. Following [13], we will also use the action $I = (1/2\pi)J$.

Quantum mechanics began in 1900 with Planck's paper [14]. He discovered that he could predict the radiation properties of black bodies provided he assumed a "quantum condition". He essentially postulated that the action variable J associated with an atom can take only the discrete values nh , where h is a universal constant and $n \in \mathbb{N}$.

An early task of quantum mechanics was to reconcile the particle and wave properties of "quantum objects" such as photons and electrons. Einstein [5] related the energy E and momentum p of a photon to the frequency ν and the wavelength λ of the corresponding wave. Noting that E/ν and $p\lambda$ are action variables (see above), he predicted that each of these equals the "minimal action" h , i. e., we have the Einstein relations

$$\begin{aligned} E &= h\nu = \hbar\omega \\ p &= h/\lambda = \hbar k \end{aligned}$$

where $\hbar = h/2\pi$. Fifteen years later de Broglie [4] proposed that these relations were valid for all particles exhibiting the wave-particle dichotomy, including the electron. It was a short step from there to finding a wave equation for which the corresponding functions (9) are solutions. This is precisely the Schrödinger equation.

In 1913 Bohr used the Planck-Einstein quantum condition to explain the spectral lines of the hydrogen atom [1]. He proposed that the electron is constrained to particular circular orbits by the quantum condition. To be more specific, he assumed that the electron had a specific energy E_m in the m -th orbit, and that it drops down (respectively jumps up) to the n -th orbit, it loses (respectively absorbs) energy $E_m - E_n$, which is carried away or brought by a photon with frequency

$$(10) \quad \omega_{m,n} = \frac{E_m - E_n}{\hbar}.$$

When Bohr used the classical Coulomb law to calculate the angular momentum L of the electron in the m -th orbit, he discovered that it was given by $L = m\hbar$ for an integer m . In fact by using the Hamiltonian theory from the next section, he and Sommerfeld showed that this coincides with Planck's quantum condition $J = m\hbar$, and the latter is also true for arbitrary closed orbital motions. Within a few years, Bohr's theory was used to predict the frequencies of the spectral lines for a variety of systems.

Bohr also formulated a fundamental asymptotic property for the spectral values which he called the *correspondence principle*. Returning to the Rydberg formula,

he observed that for large m , the electrons behaved almost classically, in the sense that one obtained overtones. More precisely, a drop of $k = m - n \ll m$ orbits resulted in the k -th overtone of a fundamental frequency $\omega_m = 2\pi Rc/m^3$:

$$\begin{aligned}\omega_{m,m-k} &= 2\pi Rc\left(-\frac{1}{m^2} + \frac{1}{(m-k)^2}\right) \\ &= 2\pi k \left(\frac{Rc}{m^3}\right) \frac{(1 - k/2m)}{(1 - 2k/m + k^2/m^2)} \\ &\sim k\omega_m.\end{aligned}$$

A similar principle applies if k is negative. We will use the notation $k \ll n$ to indicate relatively small positive or negative jumps.

In principle it would seem that we might have to consider infinitely many fundamental frequencies ω_m . However despite its nebulous character, Bohr used the correspondence principle to very accurately predict the value of the Rydberg constant R as well as the “radius” of a hydrogen atom.

Bohr’s “old” quantum theory suffered from a number of defects. In particular, the increasingly technical quantum conditions seemed unnatural, and it was difficult to calculate the “Fourier coefficients $a_{m,n}$ ”. The quantity $|a_{m,n}|^2$ measures the intensity of the frequencies $\omega_{m,n}$, or at the level of a single atom, to the probability that a jump from m to n might occur. Just as one cannot “in principle” predict when a radioactive atom might decay, one cannot say when an electron will “jump”. This is a prototypical example of the probabilistic nature of quantum mechanics.

Heisenberg concluded that the weakness of Bohr’s theory rested upon the fact that it was concerned with predicting the hypothetical singly indexed energies E_n rather than the actually observed doubly indexed frequencies $\omega_{m,n}$. As we have seen above, it was this perspective that led him to consider matrices, . However to carry out his program he had to incorporate the quantum conditions into his framework.

4. PHASE SPACE AND ACTION ANGLE VARIABLES

Quantization is typically applied to algebras of functions. Since the Hamiltonian approach to mechanics is concerned with an algebra of functions on a suitable parameter space, it is ideally suited for this process. What is particularly useful about the Hamiltonian formulation is that each function determines a one-parameter group of automorphisms, and in particular, the energy function determines the physical evolution of the system. Let us summarize this theory as quickly as possible.

Let us first suppose that we are given a parameter space $M = \mathbb{R}^n$. We let $\mathcal{D}(M)$ be the algebra of infinitely differentiable functions on M and $T(M) = M \times \mathbb{R}^n$ be the corresponding tangent space. We recall that we regard $(x, v) \in TM$ as a “tangent vector at x ”, and that it determines a corresponding directional derivative. Given $x \in M$ and $v = \sum v_j e_j \in \mathbb{R}^n$, we define

$$D_{(x,v)} : \mathcal{D}(M) \rightarrow \mathbb{R} : f \mapsto \sum v_j \frac{\partial f}{\partial x_j}(x).$$

Since tangent vectors are only used to indicate the directional derivatives that they define, we use the notation

$$(x, v) = \sum v_j \left. \frac{\partial}{\partial x_j} \right|_x$$

A *vector field* is a mapping

$$F : M \rightarrow T(M) : x \mapsto F(x) \in T_x(M),$$

and we may write

$$F(x) = \sum_{j=1}^n F_j(x) \frac{\partial}{\partial x_j}.$$

Given $f \in \mathcal{D}(M)$, we have $D_F : x \mapsto D_{F(x)}f$ is again a smooth function on M , and the mapping

$$D = D_F : \mathcal{D}(M) \rightarrow \mathcal{D}(M)$$

is a derivation of the algebra $\mathcal{D}(M)$, i.e, we have

$$D(fg) = D(f)g + fD(g).$$

As is well known, all derivations of $\mathcal{D}(M)$ arise in this manner (see [18]).

A curve

$$x : (a, b) \rightarrow M : t \mapsto x(t) = (x_1(t), \dots, x_n(t))$$

is an *integral curve* for a vector field F if for each t , $x'(t) = F(x(t))$. Thus $x(t) = (x_1(t), \dots, x_n(t))$ is just the solution to the system of first order differential equations

$$\frac{dx_j(t)}{dt} = F_j(x(t)).$$

Under appropriate conditions we may find an integral flow for the vector field, i.e, a family of mapping $\sigma_t : M \rightarrow M$ such that for each $x \in M$, $x \mapsto \sigma_t x$ is an integral curve, and furthermore $\sigma_{t+t'} = \sigma_t \circ \sigma_{t'}$, $\sigma_0 = I$. This in turn determines a one-parameter group of algebraic automorphisms α_t of the algebra \mathcal{D} , where $\alpha_t f(x) = f(\sigma_{-t}x)$. Using power series one finds a simple relationship between the derivation D_F and the automorphism group α_t :

$$\begin{aligned} D_F(f) &= \lim_{h \rightarrow 0} \frac{\alpha_h(f) - f}{h} \\ \alpha_t(f) &= e^{tD_F} f = \sum \frac{t^n}{n!} D_F^n(f) \end{aligned}$$

Turning to physics, let us consider a single oscillating particle with one degree of freedom. The Newtonian equation of motion is given by $F = ma$. Let us assume that the force F only depends upon the position q . Thus we are considering the second order equation

$$F(q(t)) = m \frac{d^2 q}{dt^2}$$

Since we have restricted to one spatial dimension, F is automatically conservative, i.e., $F(q) = -V'(q)$ for some function V , namely $V(q) = -\int F(q) dq$.

We begin by replacing Newton's equation by two first order equations. Although there are many ways this can be done (e.g. one can let $dq/dt = v$, and $dv/dt = F/m$)

Hamilton found a particularly elegant way of doing this. Specifically we use the variables q and $p = mv$. The corresponding equations are

$$(11) \quad \begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p} \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial q} \end{aligned}$$

where

$$H(q, p) = \frac{p^2}{2} + V(q).$$

We may regard the solutions curves $\gamma(t) = (q(t), p(t))$ as the integral curves of the *symplectic gradient* vector field

$$\text{sgrad}H = \frac{\partial H}{\partial p} \frac{\partial}{\partial q} - \frac{\partial H}{\partial q} \frac{\partial}{\partial p}.$$

in the *phase space* $M_2 = \mathbb{R}^2$ of variables (q, p) . This quantity is the “symplectic” analogue of the usual gradient

$$\text{grad}H = \frac{\partial H}{\partial q} \frac{\partial}{\partial q} + \frac{\partial H}{\partial p} \frac{\partial}{\partial p},$$

but it is not necessary to go into details.

In fact an arbitrary function $a(q, p)$ on M_2 determines a vector field

$$\text{sgrad}a = \frac{\partial a}{\partial p} \frac{\partial}{\partial q} - \frac{\partial a}{\partial q} \frac{\partial}{\partial p},$$

and thus corresponding flow

$$\sigma_t^a : M_2 \rightarrow M_2,$$

where $\gamma(t) = \sigma_t^a(x_0)(q(t), p(t))$ is a solution of the “Hamiltonian system”

$$\begin{aligned} \frac{dq}{dt} &= \frac{\partial a}{\partial p} \\ \frac{dp}{dt} &= -\frac{\partial a}{\partial q} \end{aligned}$$

The *Poisson brackets* of two functions a and b is defined by

$$\{a, b\} = (\text{sgrad}a)(b) = \frac{\partial a}{\partial p} \frac{\partial b}{\partial q} - \frac{\partial a}{\partial q} \frac{\partial b}{\partial p}.$$

In particular, we note that if $\{a, b\} = 0$, then letting $(q(t), p(t))$ be an integral curve of (11),

$$\frac{db}{dt} = \frac{\partial b}{\partial q} \frac{dq}{dt} + \frac{\partial b}{\partial p} \frac{dp}{dt} = \frac{\partial b}{\partial q} \frac{\partial a}{\partial p} - \frac{\partial b}{\partial p} \frac{\partial a}{\partial q} = 0,$$

i.e., the function b is constant on the orbits of a . Since $\{a, a\} = 0$, we see that a is constant on its own integral curves.

Perhaps the most striking attribute of the phase space parametrization is that the area pq of a rectangle has the dimensions

$$[\text{momentum}] \times [\text{distance}] = \mathcal{ML}^2\mathcal{T}^{-1}$$

i.e., *area is an action variable*. This link between the notion of area (or more precisely the area two-form $\Omega = dp \wedge dq$) and a physical parameter is one of the most powerful features of the Hamiltonian theory. We say that a change of variable

$Q(q, p), P(q, p)$ is *canonical* if it preserves the area form, i.e., we have that the Jacobian is 1:

$$1 = \frac{\partial(Q, P)}{\partial(q, p)} = \frac{\partial Q}{\partial q} \frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p} \frac{\partial P}{\partial q}.$$

If that is the case, then the dynamical system $Q(t) = Q(q(t), p(t))$, $P(t) = P(q(t), p(t))$ is also Hamiltonian, i.e., it has the form

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial H}{\partial Q} \\ \frac{dP}{dt} &= -\frac{\partial H}{\partial P} \end{aligned}$$

where $H(Q, P) = H(q(Q, P), p(Q, P))$. To see this we note that

$$\begin{aligned} \frac{dQ}{dt} &= \frac{\partial Q}{\partial q} \frac{dq}{dt} + \frac{\partial Q}{\partial p} \frac{dp}{dt} \\ &= \frac{\partial Q}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial Q}{\partial p} \frac{\partial H}{\partial q} \\ &= \frac{\partial Q}{\partial q} \left(\frac{\partial H}{\partial Q} \frac{\partial Q}{\partial p} + \frac{\partial H}{\partial P} \frac{\partial P}{\partial p} \right) - \frac{\partial Q}{\partial p} \left(\frac{\partial H}{\partial Q} \frac{\partial Q}{\partial q} + \frac{\partial H}{\partial P} \frac{\partial P}{\partial q} \right) \\ &= \frac{\partial H}{\partial P} \frac{\partial(Q, P)}{\partial(q, p)} = \frac{\partial H}{\partial P}, \end{aligned}$$

and a similar calculation may be used for the second equation. It is also easy to see that a canonical change of variables will leave the Poisson brackets of functions invariant. Given a Hamiltonian co-ordinate system (Q, P) , we say that Q and P are conjugate variables.

Let us assume that our system is oscillatory, i.e., all of the solution curves $(q(t), p(t))$ are closed. We may assume that $(q(0), p(0)) = (q(T), p(T))$, where T depends upon the orbit. Our goal is to find the “simplest Hamiltonian co-ordinate system” (θ, I) by using a canonical transformation, with the properties

- $H(\theta, I) = H(I)$, i.e., H doesn't depend upon θ , and
- θ increases by 2π on each closed orbit.

Given such a system, we have

$$\frac{dI}{dt} = -\frac{\partial H}{\partial \theta} = 0$$

and thus I and $H(I)$ are constant on each orbit γ . It follows that

$$\omega = \frac{d\theta}{dt} = \frac{\partial H}{\partial I}$$

is also constant on each orbit γ , i.e., $\omega = \omega(I) = \omega(\gamma)$, and $\theta(t) = \omega t + C$ for some constant C . We may assume $C = 0$, and from the second property, $\omega = 2\pi/T$.

The canonical transformation from (q, p) to (θ, I) transforms the area A enclosed by an orbit $\gamma(t) = (q(t), p(t))$ to the area R of the rectangle $0 \leq \theta \leq 2\pi$, $0 \leq I \leq I(\gamma)$. Since the purported transformation is canonical, we have

$$\oint_{\gamma} p dq = A = R = 2\pi I(\gamma),$$

where γ is the unique integral curve that passes through (q, p) . Thus assuming that we can find a canonical transformation with the desired properties, I is an action

variable. For the proof that the transformation exists (and a formula for θ) we recommend [13] or [8]. We define

$$I = \frac{1}{2\pi} \oint_{\gamma} p dq$$

to be the *action variable* and θ the *angle variable*. As one would expect, θ is multivalued since it increases by 2π on each circuit of an orbit.

The action, angle variables enable us to use Fourier series in our analysis of a periodic motion. Given an arbitrary function a on M_2 and using the action-angle variables, the function $a(I, \theta)$ will have period 2π in θ . It thus has a Fourier series

$$(12) \quad a(I, \theta) = \sum a(n) e^{in\theta}.$$

where $a(n)$ is a function of I . Substituting the solution of the Hamiltonian equations, we obtain A as a function of time:

$$a(t) = \sum a(n) e^{in\omega t}$$

where $a(n)$ is constant on the orbit.

5. THE COMMUTATION RELATION

We will identify the energy variables H and E . There is a close parallel between the classical formula

$$\omega = \frac{\partial H}{\partial I} = \frac{\partial E}{\partial I}$$

and Bohr's difference formula

$$\omega_{m,m-k} = \frac{E_m - E_{m-k}}{\hbar}$$

To make this more explicit, let us "discretize" the action variable I by letting $I = m\hbar$ and $\Delta_k I = k\hbar$. Then according to Bohr's correspondence principle, if $k \ll m$,

$$\frac{\Delta_k E}{\Delta_k I} = \frac{\Delta_k E}{k\hbar} = \frac{E_m - E_{m-k}}{\hbar} = k^{-1} \omega_{m,m-k} \sim k^{-1} k\omega = \frac{\partial E}{\partial I}.$$

It thus appears that Bohr's correspondence principle is embodied in the fact that the finite difference with respect to the discrete action variable I approximates the differential quotient with respect to the continuous action variable I . For this reason it seems justifiable to apply this to arbitrary quantum variables and their classical analogues. We will use the symbolism

$$(13) \quad \frac{\Delta_k}{\Delta_k I} \leftrightarrow \frac{\partial}{\partial I}$$

(see [18]). The difference operator will be applied to a matrix variable by the formula

$$(\Delta_k A)(m, n) = A(m, n) - A(m - k, n - k).$$

In his calculation, Heisenberg concentrated upon the Fourier coefficient functions $a(\ell)$ of a function a on the phase space M_2 and the scalar matrix coefficients $A(m, n)$ of a matrix A in the "expansions"

$$\begin{aligned} a &= \sum_{\ell} a(\ell) e^{i\ell\omega t} \\ A &= [A(m, n) e^{i\omega_{m,n} t}] \end{aligned}$$

If a is the classical function variable “reduction” of the matrix variable A , then for $\ell = m - n \ll m$ the coefficient $A(m, n)$ of $e^{i\omega_{m,n}t}$ should approximate the coefficient $a(\ell)$ of the overtone $(e^{i\omega t})^\ell$. We will write $A \rightsquigarrow a$, and $A(m, n) \rightsquigarrow a(\ell)$. We wish to show that if $A \rightsquigarrow a$ and $B \rightsquigarrow b$, then $[A, B] \rightsquigarrow \{a, b\}$.

If $j, k \ll m$ then

$$\begin{aligned} A(m, m-j) &\rightsquigarrow a(j) = \frac{1}{i} j^{-1} \frac{\partial a}{\partial \theta}(j) \text{ (for } j \neq 0) \\ (\Delta_k A)(m, m-j) &\rightsquigarrow k \hbar \frac{\partial a}{\partial I}(j). \end{aligned}$$

The equality is seen if one takes the derivative of (12) with respect to θ . The second reduction is a formal consequence of (13).

Let us suppose that we are given matrices A and B and functions a and b with $A \rightsquigarrow a$ and $B \rightsquigarrow b$. If $\ell = m - n \ll m$

$$\begin{aligned} (AB - BA)(m, n) &= \sum_{j+k=\ell} A(m, m-j) B(m-j, m-j-k) \\ &\quad - \sum_{j+k=\ell} B(m, m-k) A(m-k, m-k-j) \\ &= \sum_{j+k=\ell} [A(m, m-j) - A(m-k, m-j-k)] B(m-j, m-j-k) \\ &\quad - \sum_{j+k=\ell} A(m-k, m-j-k) [B(m, m-k) - B(m-j, m-j-k)] \\ &= \sum_{j+k=\ell} (\Delta_k A)(m, m-j) B(m-j, m-j-k) \\ &\quad - A(m-k, m-k-j) (\Delta_j B)(m, m-k) \\ &\rightsquigarrow \frac{\hbar}{i} \sum_{j+k=\ell} k \frac{\partial a}{\partial I}(j) b(k) - a(j) j \frac{\partial b}{\partial I}(k) \\ &= \frac{\hbar}{i} \sum_{j+k=\ell, k \neq 0} k \frac{\partial a}{\partial I}(j) k^{-1} \frac{\partial b}{\partial \theta}(k) - \sum_{j+k=\ell, j \neq 0} j^{-1} \frac{\partial a}{\partial \theta}(j) j \frac{\partial b}{\partial I}(k) \\ &= \frac{\hbar}{i} \left(\frac{\partial a}{\partial I} \frac{\partial b}{\partial \theta} - \frac{\partial a}{\partial \theta} \frac{\partial b}{\partial I} \right) (\ell) \\ &= \frac{\hbar}{i} \{a, b\} (\ell). \end{aligned}$$

(see (5) — we note that $\frac{\partial b}{\partial \theta}(0) = \frac{\partial a}{\partial \theta}(0) = 0$).

As Heisenberg points out in a footnote, this calculation is problematical even as a heuristic guide. Although $n - m = \ell = j + k$ is assumed “relatively small” with respect to m and n , we are summing over arbitrary j, k with $j + k = \ell$. Heisenberg explains this away by pointing out that if j is large it will follow that k is large (usually with opposite sign) and vice versa, and thus all the matrix positions $(m, m-j)$, $(m-j, m-j-k)$, $(m, m-k)$, and $(m-k, m-k-j)$ will be distant from the diagonal. He states that the corresponding matrix elements must be negligible “since they correspond to high harmonics in the classical theory”.

We conclude

$$[A, B] \rightsquigarrow \frac{\hbar}{i} \{a, b\}$$

Since

$$\{p, q\} = \frac{\partial p}{\partial p} \frac{\partial q}{\partial q} - \frac{\partial p}{\partial q} \frac{\partial q}{\partial p} = 1,$$

if we let P and Q be the quantized momentum matrices, i.e. $P \rightsquigarrow p$ and $Q \rightsquigarrow q$, we are led to postulate the commutation rule

$$[P, Q] = \frac{\hbar}{i} I.$$

This relation is the most essential algebraic ingredient of quantum mechanical computations. The reader may find early instances of these calculations in [2].

REFERENCES

- [1] N. Bohr, On the constitution of atoms and molecules: Introduction and Part I - binding of electrons by positive nuclei, *Phil. Mag.* 26 (1913), 1-25.
- [2] M. Born, *Atomic physics*, Dover, New York, 1969 ISBN 0-486-65984-4.
- [3] A. Connes, *Noncommutative geometry*. Academic Press, Inc., San Diego, CA, 1994. xiv+661 pp. ISBN: 0-12-185860-X.
- [4] L. de Broglie, Sur la definition generale de la correspondance entre onde et mouvement, *CR Acad Sci Paris* 179, 1924
- [5] A. Einstein, On a heuristic point of view about the creation and conversion of light (English translation of title), *Ann. Phys.* 17 (1905), 132-148.
- [6] G. Emch, *Mathematical and conceptual foundations of 20th-century physics*. North-Holland Mathematics Studies, 100. Notas de Matematica [Mathematical Notes], 100. North-Holland Publishing Co., Amsterdam, 1984. x+549 pp. ISBN: 0-444-87585-9
- [7] I. Gelfand and S. Fomin, *Calculus of variations*. Revised English edition translated and edited by Richard A. Silverman, Prentice-Hall, Inc., Englewood Cliffs, N.J. 1963 vii+232 pp. ISBN 0-486-41448-5 (pbk).
- [8] H. Goldstein, *Classical Mechanics*, Addison Wesley, 1950. 399 pp. ISBN 0-201-02510-8.
- [9] W. Heisenberg, Quantum-theoretical reinterpretation of kinematic and mechanical relations (translation of title), *Z. Phys.* 33, 879-893, 1925.
- [10] W. Heisenberg *Physical Principles of the Quantum Theory*, Dover, New York, 1949. 184 pp. ISBN: 486-60113-7.
- [11] Transnational College of Lex, *What is Quantum Mechanics? A Physics Adventure*, translated by J. Nambu, Language Research Foundation, Boston, 1996. 566 pp. ISBN 0-9643504-1-6.
- [12] G. Mackey, *The Mathematical Foundations of Quantum Mechanics*, W.J. Benjamin, New York, 1963.
- [13] I. Percival and D. Richards, *Introduction to dynamics*. Cambridge University Press, Cambridge-New York, 1982. v+228 pp. ISBN: 0-521-23680-0; 0-521-28149-0
- [14] M. Planck, On an improvement of Wien's equation for the spectrum (translation of title), *Verhandlungen der Deutschen Physik Gesells* 2, 202-204.
- [15] A. Sommerfeld, *Müchener Berichte* (1915), 425-458.
- [16] E. Schrödinger, E. Quantization as an eigenvalue problem (translation of title), *Ann. Physik* (1926) 79, 361-376.
- [17] V. Varadarajan, *Geometry of Quantum Theory*, Springer-Verlag, New York, 1968.
- [18] F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Scott Foresman, 270 pp. 1971.
- [19] T. Wu, *Quantum Mechanics*, World Scientific, Singapore - Philadelphia, 1985, 417pp. ISBN 9971-978-47-4.